



King's Research Portal

Document Version
Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Akmal, M., Zubair, S. Z., Jochumsen, M., Kamavuako, E. N., & Khan Niazi, I. (Accepted/In press). A tensor-based method for completion of missing electromyography data. *IEEE Access*.

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

A tensor-based method for completion of missing electromyography data

Muhammad Akmal^{1,2}, Syed Zubair¹, Mads Jochumsen³, Ernest Nlandu Kamavuako⁴, Imran Khan Niazi^{3,5,6}

¹Department of Electrical Engineering, International Islamic University, Islamabad, 44000, Pakistan

²Department of Electrical Engineering, Riphah International University, Islamabad, 44000 Pakistan

³SMI, Department of Health Science and Technology, Aalborg University, Aalborg, 9100, Denmark

⁴Centre for Robotics Research, Department of Informatics, King's College London, London, WC2B 4BG, UK

⁵Centre for Chiropractic Research, New Zealand College of Chiropractic, Auckland, 1060, New Zealand

⁶Health and Rehabilitation Research Institute, AUT University, Auckland 1010, New Zealand

Corresponding author: Imran Khan Niazi (e-mail: imran.niazi@nzchiro.co.nz).

ABSTRACT This paper discusses the recovery of missing data in surface Electromyography (sEMG) signals that arise during the acquisition process. Missing values in EMG signals occur due to either disconnection of electrodes, artifacts, muscle fatigue or incapability of instruments to collect very low amplitude signals. In many real-world EMG related applications, algorithms need complete data to make accurate and correct predictions, or otherwise, the performance of prediction reduces sharply. We employ tensor factorization methods to recover unstructured and structured missing data from EMG signals. In this paper, we use first-order weighted optimization (WOPT) of PARAFAC decomposition model to recover missing data. We tested our proposed framework against Non-Negative Matrix Factorization (NMF) and Parallel Factor Analysis (PARAFAC) on simulated as well as on offline EMG signals having unstructured missing values (randomly missing data ranging from 60% to 95%) and structured missing values. In the case of structured missing data having different channels, the percentage of missing data of a channel goes up to 50% for different movements. It has been observed empirically that our proposed framework recovers the missing data with relatively much improved accuracy in terms of Relative Mean Error (up to 50% and 30 % for unstructured and structured missing data respectively) as compared to matrix factorization methods even when the portion of unstructured and structured missing data reaches up to 95% and 50%, respectively.

INDEX TERMS EMG data, Missing data, Tensor decomposition

I. INTRODUCTION

ELECTROMYOGRAPHY (EMG) is a diagnostic technique which records the electrical activity produced by contraction of muscles. The electric activity or potential is generated by the muscle cells when these cells are electrically activated. Generally, two types of EMG exist surface EMG (sEMG) and intramuscular EMG (iEMG). sEMG is the recording of electrical activity from the muscle surface (non-invasive) whereas iEMG is recorded directly within the muscle tissue. EMG signals have many applications such as upper-limb prostheses [1], [2], [3], electric wheelchairs control [4] and muscle-computer interaction [5]. In these applications, complete EMG signals without missing data are required for efficient and successful implementation. However, practically, EMG data acquisition is not lossless. During

Signal acquisition, data is lost due to many reasons such as artifacts or disconnection of electrodes with the body [6]. These missing values in the EMG signal can cause degradation in the overall performance of healthcare applications such as myoelectric pattern recognition to predict motor intention from sEMG signals [6]. Moreover, missing values also reduce the accuracy of the classification of movements for prostheses control [7]. If data is incomplete and the percentage of missing data is large, then the classification performance and statistical power of those classification methods highly degrade, which makes it important to have complete data set. To effectively estimate the missing data, proper imputation methods must be utilized. Generally, in EMG applications, missing data had either not been recovered or estimated by simply replacing it with mean values of the neighbouring data values, which proved to be

highly sub-optimal [8]. In this work, we have focused on estimating missing values using multidimensional data structure [9], [10] based upon multilinear algebra (tensors). In this paper, we aim to recover missing values in surface EMG signals by estimating the latent structure of the data. In order to estimate latent structure, we employ tensor factorization methods which produce factor matrices which are used to produce the reconstructed tensor. We further formulate a weighted version of an error function that ignores the missing values and model only the known values which improve the estimation accuracy of recovering missing data significantly.

A. RELATED WORK

Matrix and Tensor decomposition of EMG signals have been widely studied in the literature. In [11]-[15], non-negative matrix factorization (NMF) has been applied on EMG signals for various applications, e.g. recognition of gestures, to obtain information for neural control and identification of various surface EMG signals. In [16], various matrix factorization algorithms such as Principal Component Analysis (PCA), Factor Analysis (FA), Independent Component Analysis (ICA), and Non-negative Matrix Factorization (NMF) were evaluated on EMG recording. In [17], surface EMG signals are decomposed using non-negative Tensor factorization to find the features for classification purpose. In [18], NMF was employed to identify EMG finger movements to evaluate the functional status of hand so that it can assist in hand gesture recognition, prosthetics and rehabilitation applications. In [19], FastICA method is implemented for EMG signals decomposition. In [20], NMF along with different initialization techniques was applied to acquire muscle synergies which are important for generating biomechanical tasks. In [31], higher order tensor decompositions are employed on EMG signals to estimate muscle synergies.

However, so far in the literature, missing data in EMG signals has been recovered by using ensemble classifier system [8], nonlinearities interpolation approach [21], mean data imputing [6], Empirical Decomposition Mode (EMD) [22] and marginalization and conditional-mean imputation [23]. In [8], imputation and reduced-feature models were employed to perform classification in presence of missing data but the results were not promising. In [21], missing data of up to 80% was recovered. However they tested algorithm on single subject and it is also unclear whether they recovered unstructured or structured missing data. In [6], imputation was carried out using mean of data which works poorly on non-stationary EMG data. In [22], EMD fails to recover structured missing data. In [23], the main focus was on developing classification model. However they also employed a simple mean imputation method to recover missing data. In [9] and [24], tensor factorization techniques are applied on EEG signals. However, so far, EMG signals

have not been explored that way. For the first time, in this work, missing data is recovered in EMG signals with a detailed analysis in which matrix, as well as tensor factorization methods, are employed. We apply NMF for matrix factorization and, PARAFAC and CANDECOMP/PARAFAC - Weighted OPTimization (CP-WOPT) for tensor factorization. As normalized EMG data contains non-negative values; hence, for the case of matrix factorization we apply NMF, which is the unsupervised learning algorithm used for dimensionality reduction and construction of low-dimensional approximation of observed data. NMF is more suitable because other methods such as Principal Component Analysis (PCA) produce the factors which can be positive or negative. To our knowledge, tensor factorization for recovering missing data in EMG signals has not been studied yet. In this work, for the first time, we employ the tensor factorization method to recover unstructured and structured missing data in EMG signals. We apply PARAFAC and weighted optimization (WOPT) of PARAFAC model to EMG signals and recover missing data efficiently as compared to matrix factorization techniques. The novelty of this work is found in the follows:

- For the first time, missing data in EMG signals are recovered using the tensor factorization-based method.
- We compare both matrix factorization, and tensor factorization-based approaches to recover missing data in noisy simulated data and real-world EMG data to show that the tensor-based approach outperforms matrix factorization based approach.
- We address the problem of missing data in extreme cases when up to half *consecutive* EMG samples of a particular channel are missing. Our proposed framework successfully recovers the missing data even in such an extreme case.

B. NOTATIONS AND PRELIMINARIES

Tensor $\mathcal{X}_{(i,j,k,...)}$ is a multi-dimensional array which has different modes for data representation. A tensor with one mode is a one-dimensional array referred to as a vector and with two modes is known as the matrix. A tensor of third order is shown in Fig. 1, which has three dimensions having indices $i = 1, \dots, I$, $j = 1, \dots, J$ and $k = 1, \dots, K$. In this work, a tensor is represented by uppercase Blackadder ITC letter \mathcal{X} , a matrix is represented by bold italic uppercase letter \mathbf{X} , a vector is denoted by italic bold lower case letter \mathbf{x} , and a scalar is represented by italic lowercase letter x . The individual elements of n th-order tensor are represented by lowercase letters with subscripts e.g. if N -way tensor has $(I_1 \times I_2 \times \dots \times I_N)$ samples then its n -th element is denoted by $x_{i_1 i_2 \dots i_N}$.

The *Scalar product* of two tensors \mathcal{X}, \mathcal{Y} with size $I_1 \times I_2 \times \dots \times I_N$ is defined as:

$$\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_{i_1} \sum_{i_2} \dots \sum_{i_N} x_{i_1 i_2 \dots i_N} y_{i_1 i_2 \dots i_N}$$

The *Hadamard product* of two tensors \mathcal{X}, \mathcal{Y} is defined as:

$$(\mathcal{X} * \mathcal{Y})_{i_1 i_2 \dots i_N} = x_{i_1 i_2 \dots i_N} y_{i_1 i_2 \dots i_N}$$

The *Frobenius norm* of a tensor \mathcal{X} is given by:

$$\|\mathcal{X}\|_F = \sqrt{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \dots \sum_{i_N=1}^{I_N} x_{i_1 i_2 \dots i_N}^2}$$

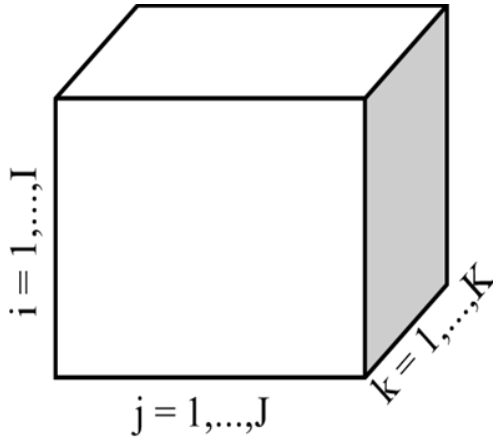


Figure 1. Tensor of 3rd order: $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$.

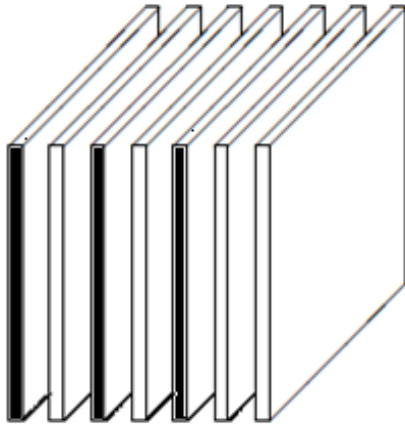


Figure 2. Missing lateral slices of a Tensor $\mathcal{X}_{j\cdot}$ (in black).

The *Weighted norm* of \mathcal{X} for two tensors \mathcal{X} and \mathcal{W}^o is defined as follows:

$$\|\mathcal{X}\|_W = \|\mathcal{W}^o * \mathcal{X}\|$$

The *Khatri-Rao product* \odot is defined as follows:

$$\mathbf{X} \odot \mathbf{Y} = [\mathbf{x}_1 \otimes \mathbf{y}_1 \quad \mathbf{x}_2 \otimes \mathbf{y}_2 \quad \dots \quad \mathbf{x}_K \otimes \mathbf{y}_K]$$

where size of matrices \mathbf{X} and \mathbf{Y} is $I \times K$ and $J \times K$ respectively. The symbol \otimes is the Kronecker product.

The *Kronecker product* \otimes is defined as follows:

$$\mathbf{X} \otimes \mathbf{Y} = \begin{pmatrix} x_{11}\mathbf{Y} & \dots & x_{1n}\mathbf{Y} \\ \vdots & \ddots & \vdots \\ x_{m1}\mathbf{Y} & \dots & x_{mn}\mathbf{Y} \end{pmatrix}$$

where \mathbf{X} is an $m \times n$ matrix and \mathbf{Y} is a $p \times q$ matrix, and the Kronecker product $\mathbf{X} \otimes \mathbf{Y}$ is the $mp \times nq$ block matrix.

The *Outer product* \circ between two vectors \mathbf{x} and \mathbf{y} is given by:

$$\mathbf{x} \circ \mathbf{y} = \mathbf{x}\mathbf{y}^T$$

where \mathbf{x} and \mathbf{y} are column vectors and their outer product gives rank-1 matrix.

Tensor mode-n unfolding, which is also called tensor matricization, is analogous to vectorizing a matrix. Mode- n unfolding of $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ re-arranges the elements of \mathcal{X} to form a matrix $\mathbf{X}_{(n)} \in \mathbb{R}^{I_n \times I_1 I_2 \dots I_{n-1} I_{n+1} \dots I_N}$, where $I_n I_{n+1} I_{n+2} \dots I_N I_1 I_2 \dots I_{n-1}$ is in a cyclic order.

The notation $[\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \dots, \mathbf{A}^{(N)}]$ defines a tensor of size $\mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ whose elements are given by:

$$([\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \dots, \mathbf{A}^{(N)}]_{i_1, i_2, \dots, i_N}) = \sum_{r=1}^R \prod_{n=1}^N a_{i_n r}^{(n)}$$

for $i_n \in \{1, \dots, I_n\}$, $n \in \{1, \dots, N\}$.

Remaining sections of the paper are organized as follows: In Section II, we explain methods which include signal processing technique and problem formulation, subjects' details, experimental setup and details of data used for evaluation. In Section III, we show results of tensor and matrix factorization methods applied on simulated and EMG data to recover both unstructured and structured missing data. Discussion on the results is given in Section IV. Section V concludes the work.

II. METHODS

A. SIGNAL PROCESSING

1) PROBLEM FORMULATION

1) NMF

The objective function for recovering missing values of the EMG data in the form of Matrix is given as:

$$f(\bar{\mathbf{X}}) = \min_{\bar{\mathbf{X}}} \|\mathbf{X} - \bar{\mathbf{X}}\|_F^2 \quad (1)$$

where $\mathbf{X} \in \mathbb{R}^{m \times n}$ is the input matrix which contains EMG data with missing values and $\bar{\mathbf{X}}$ is reconstructed matrix obtained by minimizing the objective function in (1). In order to solve (1) using NMF, the objective function in (1) becomes:

$$f(\mathbf{P}, \mathbf{Q}) = \min_{\mathbf{P}, \mathbf{Q}} \|\mathbf{X} - \mathbf{P}\mathbf{Q}\|_F^2 \quad (2)$$

where \mathbf{P} and \mathbf{Q} are $\mathbb{R}^{m \times k}$ and $\mathbb{R}^{k \times n}$ matrices, respectively." In order to apply NMF to multidimensional input data, we metricize it as a matrix \mathbf{X} with dimensions time \times channels. NMF decomposes the data of matrix \mathbf{X} into two matrices \mathbf{P} and \mathbf{Q} , as mentioned above. Our objective is to find factor matrices \mathbf{P} and \mathbf{Q} that minimize the objective function in (2).

2) PARAFAC

The objective function for recovering missing values of the EMG data in the form of tensors is given as:

$$f(\bar{\mathcal{X}}) = \min_{\bar{\mathcal{X}}} \|\mathcal{X} - \bar{\mathcal{X}}\|_F^2 \quad (3)$$

where $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is an order- N input tensor and assume its rank is R . \mathcal{X} contains EMG data with missing values and $\bar{\mathcal{X}}$ is the reconstructed tensor obtained by minimizing the objective function. To solve (3), a standard

tensor factorization is CANDECOMP/PARAFAC (CP), which can be used to find the reconstructed tensor, then the objective function in (3) becomes:

$$f(\mathbf{A}^{(1)}\mathbf{A}^{(2)}, \dots, \mathbf{A}^{(N)}) = \min_{\mathbf{A}^{(1)}\mathbf{A}^{(2)}, \dots, \mathbf{A}^{(N)}} \frac{1}{2} \left\| (\mathcal{X} - \llbracket \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \dots, \mathbf{A}^{(N)} \rrbracket) \right\|_F^2 \quad (4)$$

where $\mathbf{A}^{(n)}$ is factor matrix corresponding to n -th dimension, $\llbracket \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \dots, \mathbf{A}^{(N)} \rrbracket$ makes an order- N tensor equivalent to:

$$\overline{\mathcal{X}} \approx \llbracket \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \dots, \mathbf{A}^{(N)} \rrbracket \equiv \sum_{r=1}^R \mathbf{a}_r^{(1)} \circ \mathbf{a}_r^{(2)} \circ \dots \circ \mathbf{a}_r^{(N)} \quad (5)$$

where $\mathbf{a}_r^{(n)}$ is r -th column vector of $\mathbf{A}^{(n)}$ factor matrix, and $n = 1, 2, \dots, N$. The sum of the outer products of vectors $\mathbf{a}_r^{(n)}$ in (5) shows the CP decomposition as a sum of R rank-1 tensors to estimate a tensor. We use this CP decomposition [9] to find the factor matrices of the input tensor. The tensor in (5) is an approximation proposed by CP/PARAFAC method which is one of the standard methods for tensor factorization. In (5), a particular constraint is the value of R which is determined heuristically. We further modify this CP tensor factorization method to a weighted CP model which caters for the missing data recovery. Elementwise, (5) can be written as:

$$\bar{x}_{i_1 i_2 \dots i_N} = \sum_{r=1}^R a_{i_1 r} a_{i_2 r} \dots a_{i_N r} \quad (6)$$

for $i_1 = 1, 2, \dots, I_1, i_2 = 1, 2, \dots, I_2, \dots, i_N = 1, 2, \dots, I_N$

In mode- n unfolded (matrix) form, (5) is represented as:

$$\bar{\mathbf{X}}_{(n)} = \mathbf{A}^{(n)} (\mathbf{A}^{(-n)})^T \quad (7)$$

where

$$\mathbf{A}^{(-n)} = \mathbf{A}^{(N)} \otimes \dots \otimes \mathbf{A}^{(n+1)} \otimes \mathbf{A}^{(n-1)} \otimes \dots \otimes \mathbf{A}^{(1)}$$

In unfolded form, our objective function to find mode- n factor matrices becomes:

$$f(\mathbf{A}^{(1)}\mathbf{A}^{(2)}, \dots, \mathbf{A}^{(N)}) = \min_{\mathbf{A}^{(n)}} \frac{1}{2} \left\| (\mathbf{X}_{(n)} - \mathbf{A}^{(n)} (\mathbf{A}^{(N)} \otimes \dots \otimes \mathbf{A}^{(n+1)} \otimes \mathbf{A}^{(n-1)} \otimes \dots \otimes \mathbf{A}^{(1)})^T) \right\|_F^2$$

There exist many methods to compute CP decomposition to find a good approximation of original data such as alternating least squares (ALS) [25], [30], gradient descent (GD) [30] and enhanced line search [30] etc.

Our experiments show that conventional method such as CP decomposition only give comparable results to that of matrix factorization methods that even worsens when large amount of data is missing. To overcome this problem, we model CP factor matrices only from non-zero values of the input data. For this purpose, we multiply the input data with a weighting tensor \mathcal{W}^o with size equal to the size of input data tensor \mathcal{X} in such a way that

$$w_{i_1 i_2 \dots i_N} = \begin{cases} 1 & \text{if } x_{i_1 i_2 \dots i_N} \text{ is known} \\ 0 & \text{if } x_{i_1 i_2 \dots i_N} \text{ is missing} \end{cases}$$

for all $i_1 = 1, 2, \dots, I_1, i_2 = 1, 2, \dots, I_2, \dots, i_N = 1, 2, \dots, I_N$.

The weighted CP factorization of the EMG tensor yield factor matrices, which reconstruct the tensor using (7) to estimate the missing values.

2) CP-WOPT

CP-WOPT solves the problem of fitting the CP model to missing data by solving the following weighted least-squares objective function:

$$f_W(\mathbf{A}^{(1)}\mathbf{A}^{(2)}, \dots, \mathbf{A}^{(N)}) = \frac{1}{2} \left\| (\mathcal{X} - \llbracket \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \dots, \mathbf{A}^{(N)} \rrbracket) \mathcal{W}^o \right\|_W^2 \quad (8)$$

where \mathcal{W}^o is tensor of the same size as \mathcal{X} , and its samples are defined as:

$$w_{ijk} = \begin{cases} 1 & \text{if } x_{ijk} \text{ is known} \\ 0 & \text{if } x_{ijk} \text{ is unknown} \end{cases} \quad (9)$$

for all $i = 1, \dots, I, j = 1, \dots, J$ and $k = 1, \dots, K$.

For the sake of simplicity (8) is redefined as:

$$f_W(\mathbf{A}^{(1)}\mathbf{A}^{(2)}, \dots, \mathbf{A}^{(N)}) = \frac{1}{2} \left\| \mathcal{Y} - \mathcal{Z} \right\|^2 \quad (10)$$

where

$$\mathcal{Y} = \mathcal{W}^o * \mathcal{X} \text{ and } \mathcal{Z} = \mathcal{W}^o * \llbracket \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \dots, \mathbf{A}^{(N)} \rrbracket \quad (11)$$

The gradient equation for the weighted case would be:

$$\frac{\partial f_W}{\partial \mathbf{A}^{(n)}} = (\mathbf{Z}^{(n)} - \mathbf{Y}^{(n)}) \mathbf{A}^{(-n)}, \quad (12)$$

for $n = 1, \dots, N$.

Our main objective is to find factor matrices $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R}$ for $n = 1, \dots, N$ that minimize the weighted objective function in (8). Once gradients in (12) are known, then any gradient-based optimization method can be used to solve the optimization problem. We use CP-WOPT [25] and the nonlinear conjugate gradient (NCG) as the optimization method with Hestenes-Stiefel updates [26]. The stopping conditions of both tensor based algorithms were based on the relative change in the function value f_W in (8) (set to 10^{-8}). The maximum number of iteration is set to 10^3 and the maximum number of function evaluations is set to 10^4 . These choices are based on the values used in [9]. The brief methodology of CP-WOPT is summarized below:

Algorithm Methodology of CP-WOPT

Task: To find gradient matrices $\mathbf{G}^{(n)}$ that minimize the weighted objective function in (6).

Input: \mathcal{X} (Input tensor with missing values)

Output: $\mathbf{G}^{(n)}$

Steps to compute $\mathbf{G}^{(n)}$:

1. Compute $\mathcal{Y} = \mathcal{W}^o * \mathcal{X}$
2. Compute $\mathcal{Z} = \mathcal{W}^o * \llbracket \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \dots, \mathbf{A}^{(N)} \rrbracket$
3. Compute value of functions: $f = \frac{1}{2} - \langle \mathcal{Y}, \mathcal{Z} \rangle + \frac{1}{2} \|\mathcal{Z}\|^2$
4. Compute $\mathcal{T} = \mathcal{Y} - \mathcal{Z}$

Repeat for $n = 1, \dots, N$:

5. $\mathbf{G}^{(n)} = -\mathbf{T}_{(n)} \mathbf{A}^{(-n)}$

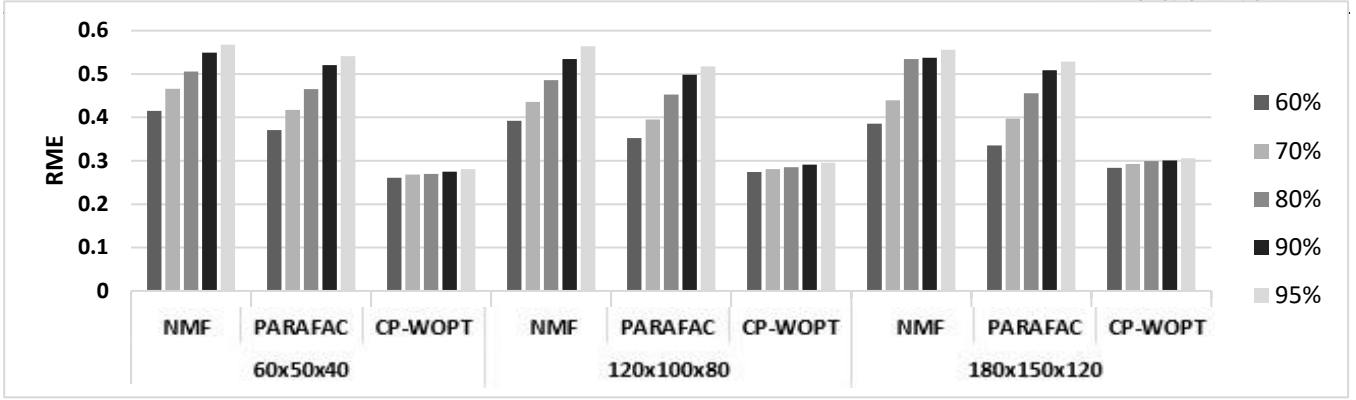


Figure 3. RME of NMF, PARAFAC and CP-WOPT methods for 60%, 70%, 80%, 90% and 95% missing data for synthetic data of sizes 60x40x40, 120x100x80 and 180x150x120.

Assume $\mathcal{Y} = \mathcal{W}^o * \mathcal{X}$ is pre-computed as both \mathcal{W}^o and \mathcal{X} remain the same in the algorithm. The gradient is computed as a series of matrices $\mathbf{G}^{(n)} \equiv \frac{\partial f_W}{\partial \mathbf{A}^{(n)}}$ for $n = 1, \dots, N$. While $\mathbf{T}_{(n)}$ is the unfolding of the tensor \mathcal{T} in mode n . Once gradients $\mathbf{G}^{(n)}$ are computed, then any gradient-based optimization method can be used to solve the optimization problem.

B. SUBJECTS DETAIL

For this research, we have used sEMG data used by M. Zia ur Rehman *et al.* in [1] and A Waris *et al.* in [27]. Ten subjects (all male) were recruited for EMG data acquisition. Ages of all subjects ranged from 18 to 38 years old (mean \pm standard deviation (SD), 24.5 \pm 2.3y). All subjects were healthy with no neuromuscular disorders. The procedures were in accordance with the Declaration of Helsinki and approved by the local ethical committee of Riphah International University (approval no: ref# Riphah/RCRS/REC/000121/20012016). Subjects provided written informed consent prior to the experimental procedures.

C. THEORETICAL FRAMEWORK

Missing data in EMG has been categorized into two types: 1) unstructured missing data 2) structured missing data. If the observed data in the original structure is missing randomly, then such a pattern of missing data is categorized as unstructured missing data. For example, samples of EMG data missing at random entries. However, if the data

is missing in some consistent and structured way, it is termed as structured missing data. For example, 25% consecutive values of an EMG channel are missing either at the start, middle or end of data acquisition process/session. This block of missing values is repeated randomly in other channels of EMG data.

D. EXPERIMENTAL SETUP

Surface EMG signals were acquired using six surface EMG electrodes. Three electrodes were placed on flexor and three electrodes on extensor muscles. The sampling frequency of

surface EMG signals was 8 kHz, whereas we filtered it using bandpass filter of 3rd order with bandwidths 20-500 Hz. Total of four-hand motions was performed by each subject: (1) hand open (2) hand close (3) pronation and (4) extend a hand. For each session, each hand motion was repeated four times with a contraction and relaxation time of five seconds, and hence a single session took a time of 400 seconds.

E. DATA ANALYSIS

We applied NMF, PARAFAC and CP-WOPT on simulated and EMG data to recover both types of missing data. In order to carry out the comparison, we assess the performance of methods based on validation metric termed as Relative Mean Error (RME) mentioned in (14).

1) SYNTHETIC DATA

We generated tensor of size $\mathbb{R}^{I \times J \times K}$ and kept a number of true factors $R = 5$. In order to test the performance of different methods to recover unstructured missing data from synthetic data, we produced synthetic data of different size such as $60 \times 50 \times 40$, $120 \times 100 \times 80$, $180 \times 150 \times 120$. For the case of structured missing data, we test the methods on a dataset of size $120 \times 100 \times 80$. Factor matrices \mathbf{A} , \mathbf{B} and \mathbf{C} were generated with sizes: $\mathbb{R}^{I \times R}$, $\mathbb{R}^{J \times R}$ and $\mathbb{R}^{K \times R}$ respectively. All the factor matrices were randomly chosen from $\mathcal{N}(0,1)$ and then normalized every column to unit length. ϵ then create the data tensor as:

$$\mathcal{X} = [\mathbf{A}, \mathbf{B}, \mathbf{C}] + \eta \frac{[\mathcal{X}]}{[\mathcal{N}]} \quad (13)$$

Here \mathcal{N} is a noise tensor (of the same size as \mathcal{X}) in which all samples were drawn from Gaussian i.i.d. distribution with mean zero and variance one. The term $[\mathbf{A}, \mathbf{B}, \mathbf{C}]$ is a

tensor being constructed from factor matrices \mathbf{A} , \mathbf{B} and \mathbf{C} where η is noise parameter which has value 0.1.

In order to implement matrix and tensor-based factorization methods, we set some samples of a tensor to zero to model missing data. In the case of weighted tensor factorization, the tensor \mathcal{W}^o indicates the binary values zero or one where zero and one represent missing and known values, respectively. In particular, we have considered two cases of missing data: (1) Unstructured missing data and (2) Structured missing data. In case of unstructured missing data, we randomly set some percentage of data (from 60% to 95% of total data) to zero in the tensor \mathcal{W}^o whereas in case of structured missing

data we set, within multiple channels, large consecutive-samples (up to 50%) to zero which is usually the case in practical situations. It is tantamount to a situation where half EMG data of multiple channels is missing for a particular movement.

2) EMG DATA

For EMG data acquisition, six electrodes were used to collect EMG signals on a single day. The movement-wise size of data was $320000 \times 6 \times 4$, which was down-sampled to $80000 \times 6 \times 4$. 80000 is the number of samples, 6 represents total number of electrodes/channels and 4 is total number of movements for which EMG data was collected. After downsampling EMG data, we normalized it between 0 and 1. Surface EMG data in the form of a tensor \mathcal{X} can be viewed as $\mathcal{X} \in \mathbb{R}^{80000 \times 6 \times 4}$ for each of four movements. If we relate it with Fig. 1, then $I = 80000$, $J = 8$ and $K = 4$ where I , J and K represent samples of EMG data, total number of channels and total number of movements respectively. Fig. 2 shows lateral slices of a tensor $\mathcal{X}_{:,j}$ which in our case represents EMG data of channels. As we have total of six channels, hence $\mathcal{X}_{:,1}$ would be a slice representing EMG data of first channel and so on. The black slices in Fig. 2 are the ones with missing EMG data.

We removed samples in the following two ways: 1) unstructured and 2) structured missing data. Fig. 5(b) shows unstructured missing data in which individual samples of EMG data of Fig. 5(a) are randomly missing. EMG data for four movements is shown in Fig. 6(a) where higher amplitudes show intervals of hand movements and lower amplitudes represent intervals of rest. Fig. 6(b) shows structured missing data in which first half part (chunk of consecutive samples) is removed whereas Fig. 6(d) shows structured missing data in which second half part is removed. In our experiments, we recover these missing intervals in extreme cases where half consecutive EMG samples of a particular channel are missing. However in [9] and [32], it is known that missing data cannot be recovered by low rank tensor completion if entire slice is completely missing.

Matrix and tensor factorization methods are applicable to a wide range of real world signals and do not depend on statistical and mathematical properties of the signals. However initialization of matrix and tensor factorization models does have effect on estimation of the signals in question. For example, in our case, EMG is stochastic in nature hence initialization of our matrix and tensor factorization models with random values help to recover the missing values efficiently. Moreover our proposed framework can be employed on other biomedical signals as well e.g. EEG signals.

F. EVALUATION METRIC

Let \mathbf{X} be the original data and let $\hat{\mathbf{X}}$ be the estimated data produced by the matrix or tensor factorization methods.

Then the Relative Mean Error (RME) is:

$$\text{RME} = \frac{\|\mathbf{X} - \hat{\mathbf{X}}\|_F}{\|\mathbf{X}\|_F} \quad (14)$$

The best possible score is zero which shows the recovered data matches with original data completely.

G. SIMULATION ENVIRONMENT

We used Matlab 2017a on Windows 8 operating system with a core i3 processor and 6 GB RAM. CP-WOPT is implemented using Tensor Toolbox.

I. STATISTICS

A three-way ANOVA was used to assess which method had the least amount of RME. Three factors: methods (NMF, PARAFAC and CP-WOPT), Movement type (hand open, hand close, pronation and extend hand) and missing data percentage (10%, 20%, 30%, 40% and 50%) were used, post hoc pairwise comparisons were made using Tukey's HSD tests if required. Statistical significance was set at $P < 0.05$ for all comparisons.

III. RESULTS

In this paper, the proposed framework is tested on both synthetic and EMG data set to recover both unstructured and structured missing data. For both cases, we assess the performance of our proposed framework CP-WOPT against matrix-based method NMF and tensor-based method PARAFAC to recover both types of missing data. Our results show that missing values can be efficiently recovered with CP-WOPT as compared to NMF and PARAFAC.

A. Estimation performance on synthetic data

In Fig. 3, we compare the estimation performance of matrix and tensor-based factorization methods to recover unstructured missing data in the synthetic dataset for different proportions, e.g. 60%, 70%, 80%, 90% and 95%. In Fig. 4, we show the capability of different methods to recover structured missing data. Structured missing data is modelled by replacing entire 10,20,30,40 and 50 columns (which are channels in case of EMG data) with zeroes.

B. Estimation performance on real EMG data

Fig. 5 shows a segment of original EMG data with no missing values, the same EMG segment with unstructured missing values, and lastly the recovered EMG signal. A segment of the original EMG signal is shown in Fig. 5(a) with no missing values and it contains information of movement of muscle from a single channel. Fig. 5(b) shows the same EMG signal with unstructured missing values, which are the input signal to factorization methods. It can be seen in Fig. 5(b) that a lot of values with different amplitudes are replaced by zeroes to model unstructured missing data. Fig. 5(c) shows a recovered EMG signal when CP-WOPT is applied on the EMG signal of Fig. 5(b). It can be seen in Fig. 5(c) that all

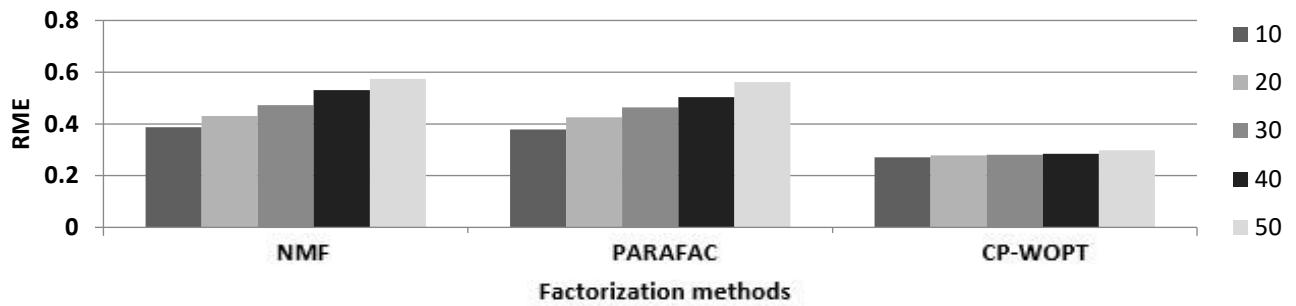


Figure 4. RME of NMF, PARAFAC and CP-WOPT methods for 10, 20, 30, 40 and 50 columns missing in structured manner from synthetic data of size 120x100x80.

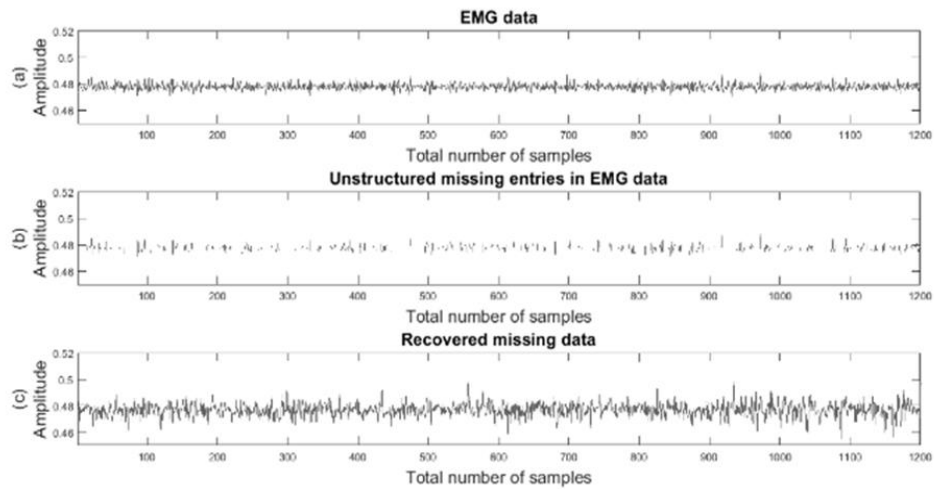


Figure 5. (a) Original EMG data (b) Unstructured missing data (c) Recovered missing data by CP-WOPT.

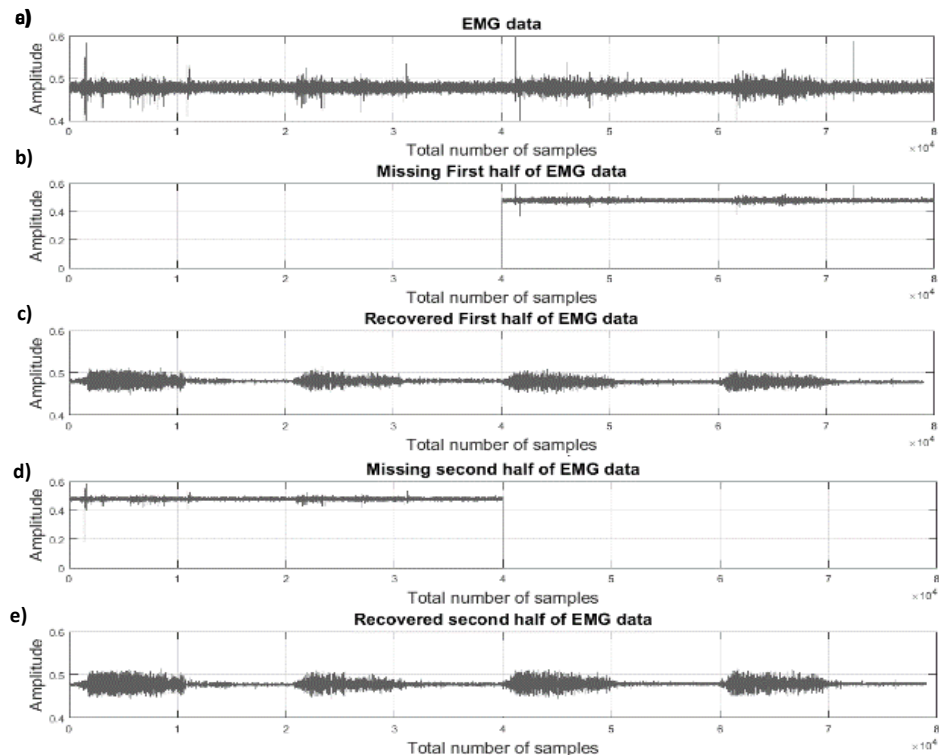


Figure 6. (a) Original EMG data (b & d) First and Second half of a channel missing (c) & (e) Recovered missing channel by CP-WOPT.

the missing values that were replaced by zeroes were successfully recovered with amplitudes around 0.48.

Fig. 6 illustrates a segment of EMG data with no missing values having same four movements where each movement exists at higher amplitudes from which first half (with two movements) and second half (with two movements) is removed and then recovered. In Fig. 6(a), EMG signal with no missing values is shown that has been obtained from a particular channel. The four epochs of higher amplitudes indicate execution of movement however it can be seen that there is a very small difference between amplitudes of movement and no-movement (at rest) epochs. In Fig. 6 (b & d), first and second half (the worst case of removing 50% of data) of channel values is removed to model the structured missing data. It is tantamount to the scenario where data of two movements is missed completely. In Fig. 6(c & e), recovered signal by CP-WOPT is shown in which it can be seen that the difference between amplitudes of movement and no-movement epochs have increased which clearly differentiate epochs.

Fig. 7 shows a comparison of three methods to recover unstructured missing data. There was a significant decrease ($P < 0.05$) in the RME value with CP-WOPT as compared to PARAFAC and NMF across all four movements and different percentage of missing data. From each of four movements, we remove 60%, 70%, 80%, 90% and 95% data randomly in an unstructured manner.

In Fig. 8, results are shown when NMF, PARAFAC and CP-WOPT are applied, respectively, to recover structured missing data. Results clearly show that CP-WOPT outperformed PARAFAC and NMF in recovering structured even for the extreme case when half of the channel data is missing. In structured missing data, we gradually increased the proportion of missing data from 10% to 50%. Removing 10% data from first half means data removal of first 10% samples from all six channels of particular movement whereas removing 50% data means data removal of first 50% samples (as shown in Fig. 6(b)) from all channels. Likewise, removing 10% data from second half means data removal of last 10% of samples from all six channels of particular movement whereas removing 50% means data removal of last 50% of samples (as shown in Fig. 6(d)).

In Fig. 9, we show computational complexity of NMF, PARAFAC and CP-WOPT. It can be seen that CP-WOPT takes slightly more time than NMF and PARAFAC to estimate 10%, 20%, 30%, 40% and 50% structured missing values to produce the reconstructed EMG data.

IV. DISCUSSION

We assessed matrix and tensor factorization techniques to evaluate their performance to recover missing data for synthetic and real EMG data. For matrix and tensor factorization we applied NMF, and PARAFAC and CP-WOPT respectively. One of the reason for tensor factorization to outperform NMF is the arrangement of EMG data in a multidimensional way. This multidimensional arrangement of the data to constitute a tensor captures the global structure of observed data and models it efficiently by

covering entire spatial and temporal dimension with an additional feature of multi-mode correlations. Moreover CP-WOPT outperforms PARAFAC as well because it is a weighted version of PARAFAC and models only the known values of EMG data. The key finding is that tensor factorization technique CP-WOPT in which only known samples are modelled outperformed both NMF and PARAFAC. The performance of NMF and PARAFAC to recover missing data was almost the same as both the methods model, both known and unknown values.

Although PARAFAC is a tensor-based technique with the benefit of preserving the multi-way nature of data, yet its performance is comparable with NMF. The results reveal that CP-WOPT outperformed both NMF and PARAFAC to recover both unstructured and structured missing data. Usually, factorization methods find latent factors and then exploits those latent factors to predict the missing values. However, Matrix factorization based latent factors only capture two-dimensional linear relationships for estimating missing values, which can be improved if multi-linear relations are used. The main advantage of working through latent factors is that they let us take into account the information of the tensor explicitly by exploiting the multilinear interactions between obtained latent factors. For example, in our case, EMG data has dimensions: samples \times channels \times movements.

Once we obtain latent factors, the inter-relation between factors of EMG data in each mode can be analyzed, such that columns of the first factor explicitly describes EMG signal, columns of seconds factor describes channels and columns of the third factor depicts movement-wise data. The main advantage of employing tensor factorization is that solution provided by it is unique [28]. Moreover, tensor factorization offers better computational capabilities and storage [29].

We divided the missing data into two categories: 1) unstructured missing data and 2) structured missing data. CP-WOPT gave promising results in recovering unstructured and structured (which is a more realistic assumption in Muscle-Computer Interface) missing data. This study is a preliminary step in the feasibility of improving the accuracy of classification methods to efficiently classify hand movements using surface EMG signals. Performance of classification methods will improve because firstly missing data is replaced with efficiently calculated estimated data and secondly; it increases the total size of data. However, the study presented here is an offline analysis and based on a small number of able-bodied subjects, which limits the possibility of generalizing the results. Furthermore, the relation between RME and classification performance needs to be developed so that it can be claimed that improved RME improves the classification performance for myoelectric control application.

V. CONCLUSION

In this paper, we addressed the problem of recovering two types of missing data in surface EMG signals: unstructured

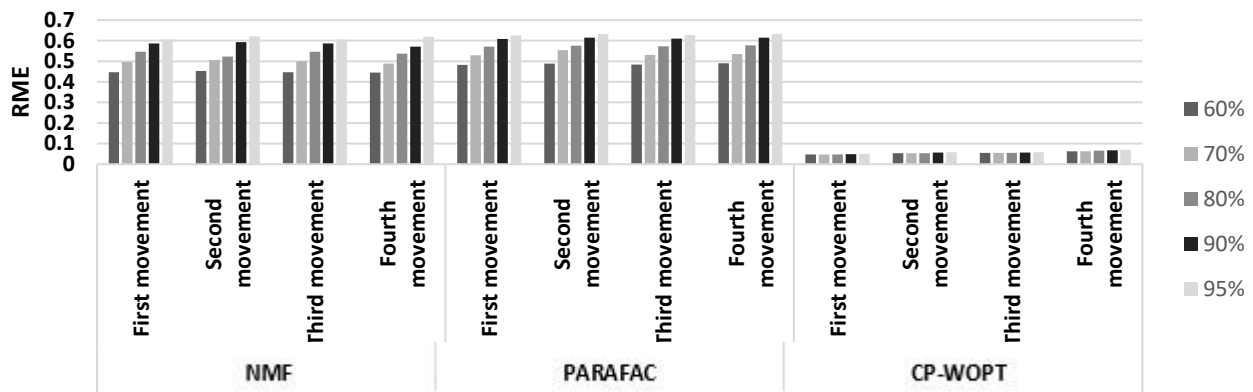


Figure 7. RME for recovering 60%, 70%, 80%, 90% and 95% unstructured EMG missing data by NMF, PARAFAC and CP-WOPT.

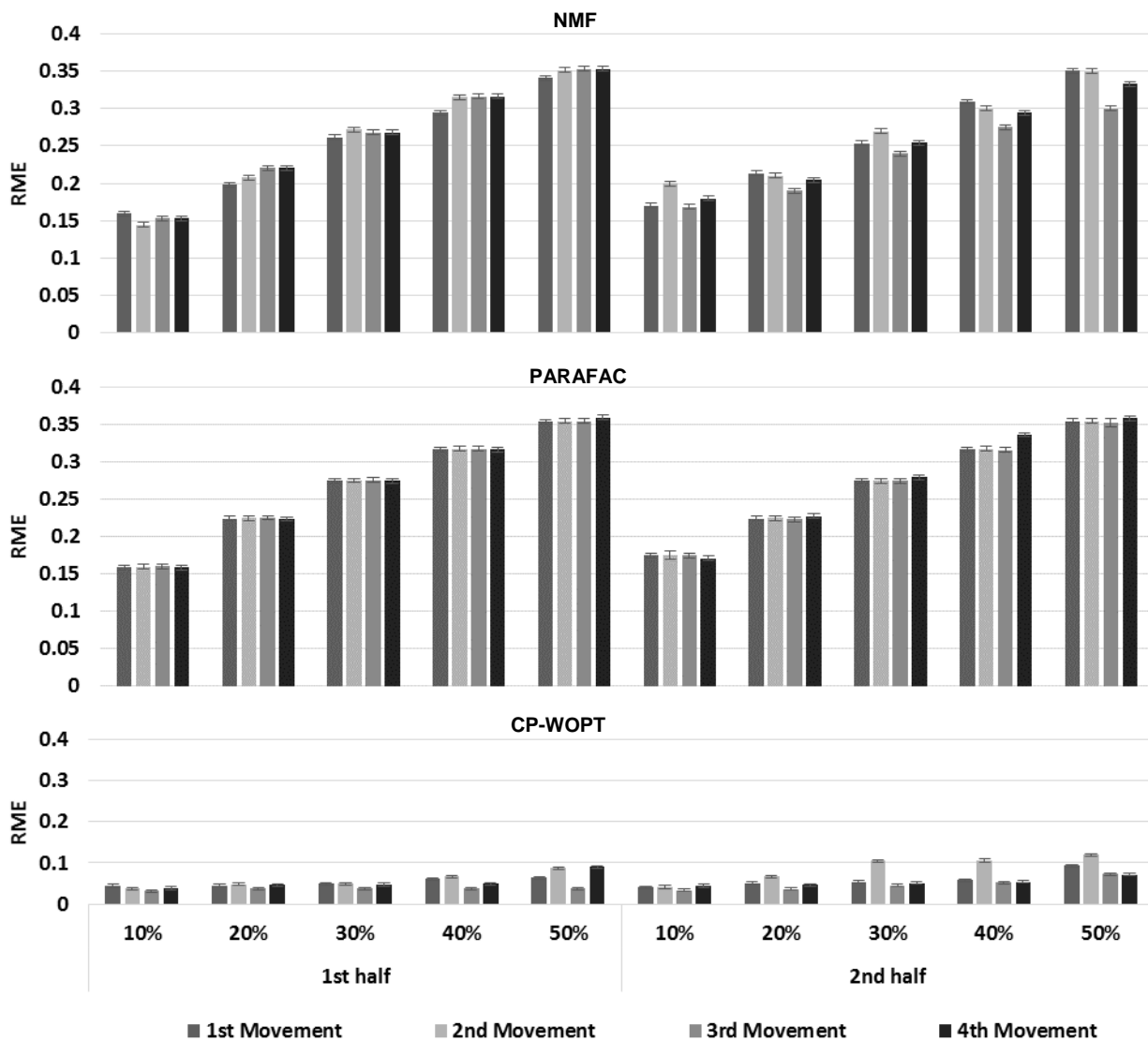


Figure 8. RME for recovering structured missing samples from first and second half of real EMG data by NMF, PARAFAC and CP-WOPT.

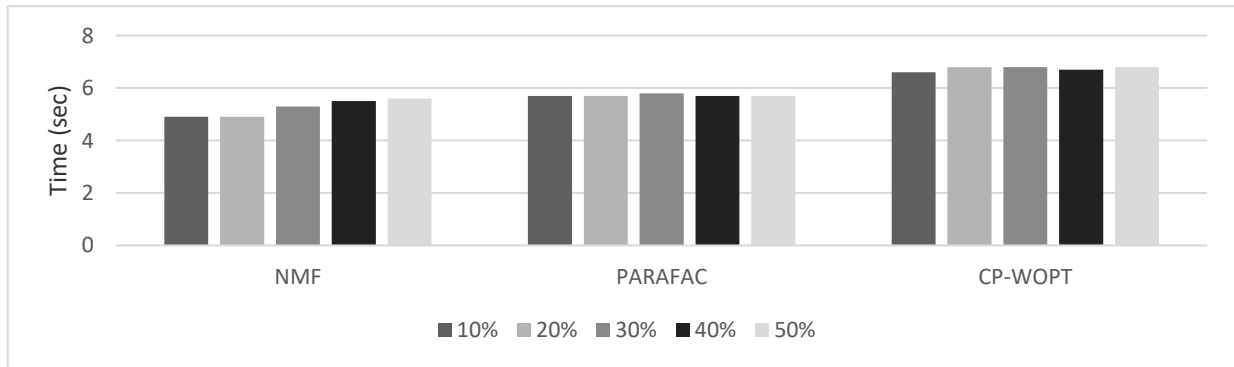


Figure 9. Comparison of computational time of NMF, PARAFAC and CP-WOPT.

and structured missing data, using NMF which is a matrix-factorization method, and PARAFAC and CP-WOPT which are tensor-factorization methods. In NMF, EMG data is matricized unlike in PARAFAC and CP-WOPT. CP-WOPT outperformed both NMF and PARAFAC in terms of RME because CP-WOPT has the ability to recover missing data efficiently such that it models only the known samples from EMG signals, which make it very useful for improving the performance of classification methods. However, this study is limited to offline analysis of sEMG signals.

REFERENCES

1. M. Zia urRehman, S. Gilani, A. Waris, I. Niazi, G. Slabaugh, D. Farina and E. Kamavuako, "Stacked Sparse Autoencoders for EMG-Based Classification of Hand Motions: A Comparative Multi Day Analyses between Surface and Intramuscular EMG", *Applied Sciences*, vol. 8, no. 7, p. 1126, 2018.
2. M. Zia urRehman, A. Waris, S. Gilani, M. Jochumsen, I. Niazi, M. Jamil, D. Farina and E. Kamavuako, "Multiday EMG-Based Classification of Hand Motions with Deep Learning Techniques", *Sensors*, vol. 18, no. 8, p. 2497, 2018.
3. E. Scheme and K. Englehart, "Electromyogram pattern recognition for control of powered upper-limb prostheses: State of the art and challenges for clinical use", *The Journal of Rehabilitation Research and Development*, vol. 48, no. 6, p. 643, 2011.
4. A. Phinyomark, C. Limsakul and P. Phukpattaranont, "A Review of Control Methods for Electric Power Wheelchairs Based on Electromyography Signals with Special Emphasis on Pattern Recognition", *IETE Technical Review*, vol. 28, no. 4, p. 316, 2011.
5. T. S. Saponas, D. S. Tan, D. Morris, R. Balakrishnan, J. Turner, and J. A. Landay, "Enabling always-available input with muscle-computer interfaces," in *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, 2009, pp. 167-176.
6. A. NABER, "Stationary Wavelet Processing and Data Imputing in Myoelectric Pattern Recognition on an Embedded System."
7. D. Farina, N. Jiang, H. Rehbaum, A. Holobar, B. Graimann, H. Dietl, *et al.*, "The extraction of neural information from the surface EMG for the control of upper-limb prostheses: emerging avenues and challenges," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, pp. 797-809, 2014.
8. C. Setz, J. Schumm, C. Lorenz, B. Arnrich, and G. Tröster, "Using ensemble classifier systems for handling missing data in emotion recognition from physiology: one step towards a practical system," in *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, 2009, pp. 1-8.
9. E. Acar, D. Dunlavy, T. Kolda and M. Mørup, "Scalable tensor factorizations for incomplete data", *Chemometrics and Intelligent Laboratory Systems*, vol. 106, no. 1, pp. 41-56, 2011.
10. A. Waris, I. K. Niazi, M. Jamil, O. Gilani, K. Englehart, W. Jensen, *et al.*, "The effect of time on emg classification of hand motions in able-bodied and transradial amputees," *Journal of Electromyography and Kinesiology*, vol. 40, pp. 72-80, 2018.
11. J. Yousefi and A. Hamilton-Wright, "Characterizing EMG data using machine-learning tools", *Computers in Biology and Medicine*, vol. 51, pp. 1-13, 2014.
12. Ning Jiang, K. Englehart and P. Parker, "Extracting Simultaneous and Proportional Neural Control Information for Multiple-DOF Prostheses From the Surface Electromyographic Signal", *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 4, pp. 1070-1080, 2009.
13. F. Theis and G. García, "On the use of sparse signal decomposition in the analysis of multi-channel surface electromyograms", *Signal Processing*, vol. 86, no. 3, pp. 603-623, 2006.
14. Y. Yazama, Y. Mitsukura, M. Fukumi, and N. Akamatsu, "Recognition from EMG signals by an evolutionary method and non-negative matrix factorization," in *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, 2003, pp. 594-600.
15. A. Ajiboye and R. Weir, "Muscle synergies as a predictive framework for the EMG patterns of new hand postures", *Journal of Neural Engineering*, vol. 6, no. 3, p. 036004, 2009.
16. M. Tresch, V. Cheung and A. d'Avella, "Matrix Factorization Algorithms for the Identification of Muscle Synergies: Evaluation on Simulated and Experimental Data Sets", *Journal of Neurophysiology*, vol. 95, no. 4, pp. 2199-2212, 2006.
17. P. Xie and Y. Song, "Multi-domain feature extraction from surface EMG signals using nonnegative tensor factorization," in *Bioinformatics and Biomedicine (BIBM), 2013 IEEE International Conference on*, 2013, pp. 322-325.
18. G. Naik and H. Nguyen, "Nonnegative Matrix Factorization for the Identification of EMG Finger Movements: Evaluation Using Matrix Analysis", *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 2, pp. 478-485, 2015.
19. M. Chen and P. Zhou, "A Novel Framework Based on FastICA for High Density Surface EMG Decomposition", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 1, pp. 117-127, 2016.
20. M. Soomro, S. Conforto, G. Giunta, S. Ranaldi and C. De Marchis, "Comparison of Initialization Techniques for the Accurate Extraction of Muscle Synergies from Myoelectric Signals via Nonnegative Matrix Factorization", *Applied Bionics and Biomechanics*, vol. 2018, pp. 1-10, 2018.
21. A. K. A Al-naqeeb, A. A Ibrahim, and Q. S Tawfeeq, "New Nonlinearities Interpolation Approach Applied to Surface EMG Signal," in *Proceedings of the International Joint Conference on Biomedical Engineering Systems and Technologies-Volume 4*, 2014, pp. 171-177.

22. T. Sidekierskiene and R. Damasevicius, "Reconstruction of missing data in synthetic time series using emd," in *CEUR Workshop Proceedings*, 2016, pp. 7-12.
23. Q. Ding, J. Han, X. Zhao and Y. Chen, "Missing-Data Classification With the Extended Full-Dimensional Gaussian Mixture Model: Applications to EMG-Based Motion Recognition", *IEEE Transactions on Industrial Electronics*, vol. 62, no. 8, pp. 4994-5005, 2015.
24. T. N. Thieu, H.-J. Yang, T. D. Vu, and S.-H. Kim, "Tensor Completion Based on Tucker Factorization for Missing Value Imputation," *International Journal of Bioscience, Biochemistry and Bioinformatics*, vol. 6, p. 34, 2016.
25. T. Kolda and B. Bader, "Tensor Decompositions and Applications", *SIAM Review*, vol. 51, no. 3, pp. 455-500, 2009.
26. S. Wright and J. Nocedal, "Numerical optimization," *Springer Science*, vol. 35, p. 7, 1999.
27. A. Waris, I. K. Niazi, M. Jamil, K. Englehart, W. Jensen, and E. N. Kamavuako, "Multiday Evaluation of Techniques for EMG Based Classification of Hand Motions," *IEEE journal of biomedical and health informatics*, 2018.
28. J. Kruskal, "Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics", *Linear Algebra and its Applications*, vol. 18, no. 2, pp. 95-138, 1977.
29. M. Mørup, "Applications of tensor (multiway array) factorizations and decompositions in data mining", *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, no. 1, pp. 24-40, 2011.
30. P. Comon, X. Luciani, and A. L. De Almeida, "Tensor decompositions, alternating least squares and other tales," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 23, pp. 393-405, 2009.
31. Ebied, A., Kinney-Lang, E., Spyrou, L. and Escudero, J. (2019). Muscle Activity Analysis Using Higher-Order Tensor Decomposition: Application to Muscle Synergy Extraction. *IEEE Access*, 7, pp.27257-27271.
32. Ying, J., Lu, H., Wei, Q., Cai, J., Guo, D., Wu, J., Chen, Z. and Qu, X. (2017). Hankel Matrix Nuclear Norm Regularized Tensor Completion for N -dimensional Exponential Signals. *IEEE Transactions on Signal Processing*, 65(14), pp.3702-3717.